2017

Representation of Positive Alpha-Stable Network Traffic Through Levy Mixtures

Bollmann, Chad; Tummala, Murali; McEachen, John

Bollmann, Chad, Murali Tummala, and John McEachen. "Representation of positive alpha-stable network traffic through levy mixtures." Signals, Systems, and Computers, 2017 51st Asilomar Conference on. IEEE, 2017.
http://hdl.handle.net/10945/60771

Representation of Positive Alpha-Stable Network Traffic Through Levy Mixtures

Chad Bollmann
Department of Electrical
and Computer Engineering
Naval Postgraduate School
Monterey, California
Email: cabollma@nps.edu

Murali Tummala
Department of Electrical
and Computer Engineering
Naval Postgraduate School
Monterey, California
Email: mtummala@nps.edu

John McEachen
Department of Electrical
and Computer Engineering
Naval Postgraduate School
Monterey, California
Email: mceachen@nps.edu

***Abstract*—Aspects of network traffic, among other impulsive time series, can be more accurately represented using the family of stable distributions. Simple, closed form solutions for stable distributions do not exist, other than special cases. Mixtures of one of these special cases, the Lévy (or Pearson V) distribution, can be used to provide a closed-form approximation of positive α-stable (PαS) distributions. We show that for a specific network traffic trace, accurate closed-form approximations of a PαS time series can be obtained with only four mixture components. Additionally, we provide an algorithm for creating Lévy Mixture Approximations (LMAs) and demonstrate that non-linear methods can improve model accuracy while constraining the number of components and computational cost. This approach provides a computationally-tractable, accurate model for non-Gaussian, positive (or negative) time series such as network traffic. This model is in a form that is less costly for follow-on processing and detection, potentially facilitating real-time applications.**

I. Introduction

Aspects of network traffic have been known to exhibit non-Poisson (and non-Gaussian) behavior for some time [1], [2], [3]. Depending on the specific feature, or attribute, of network traffic that is monitored, as well as the window over which the traffic statistics are collected, the resulting distribution will frequently be wholly-positive, non-symmetric, and heavy-tailed. These are three characteristics that a Gaussian distribution can only approximate with error. However, Gaussian-based modeling and detection continues to be the *de-facto* implementation in many statistical and machine learning anomaly detection methods.

It has been demonstrated that the α-stable distribution can provide a more accurate, and flexible, fit of network traffic than traditionally-used Gaussian distributions and their alternatives for a range of network traffic aggregations [4], [5]. Many fields have used α-stable models to improve modeling accuracy [6], though their adoption has likely been limited by the computational inefficiencies and complexities that result from the lack of closed-form solutions for (non-special case) stable distributions [7]. Previous work in detection has repeatedly demonstrated that applying α-stable estimation and detection algorithms when encountering non-Gaussian environments can significantly improve detector performance [8], [9].

This work was funded through a grant from the Laboratory for Telecommunications Sciences.

Our work seeks to develop a computationally-tractable method that improves the modeling accuracy of network traffic features characterized by α-stable distributions. This modeling method will ultimately serve as the basis of a non-Gaussian detector with the goal of improving detection accuracy over Gaussian-based implementations.

The objective of this paper is to present results demonstrating that appropriately-distributed network traffic time series can be accurately modeled using a finite mixture of closed-form Lévy distributions [10]. To our knowledge, this approach has not been applied in the literature to real-world data. We develop an algorithmic reformulation of the existing work to facilitate implementation, then use our algorithm to demonstrate that good accuracy can be obtained with very few components and minimal post-processing optimization.

This paper is organized as follows. Section II and III provide background on network traffic modeling and alpha-stable distributions. Section IV describes PαS approximation using Lévy distribution and our algorithm. Section V conveys additional results. Conclusions are discussed in Section VI.

II. Modeling Network Traffic

In this section we briefly discuss historical traffic modeling, and the limitations of these methods that require a new approach to the underlying assumptions used to develop traffic models and anomaly detectors.

The Poisson and Gaussian distributions were among the first used to model network traffic, and their limitations have long been identified [11]. Common features of network traffic such as byte and packet count, connection time, *etc.* are asymmetric and heavy-tailed [1].

In an attempt to reflect characteristics of network traffic not suited to these distributions (such as high variance and self-similarity) alternative models were examined with varying success, including the Weibull, Pareto, Gaussian, and gamma [3]. All of these models, including our proposed model, incur some measure of error when approximating the underlying time series. But it is our belief that the flexibility of the α-stable model allows the most *accurate reflection* of fundamen-

tal characteristics of network traffic, and will thus provide the *closest approximation* of many features of network traffic that are suitable for anomaly detection [5].

We will now review the aspects of this distribution that are most pertinent to our work.

III. The α-Stable Family

Stable distributions were originally developed in the 1920s by Lévy and Khinchine [12], and were first popularly applied to financial analysis and forecasting in the 1960s by Mandelbrot [7]. The family of stable distributions is also referred to in the literature as Lévy stable, Pareto stable (or Paretian), and α-stable. Stable distributions have been applied to improve the accuracy of modeling random processes that exhibit non-Gaussian behavior in a significant body of work, across fields as varied as finance, signal processing (including radar, sonar, and wireless noise), animal behavior, and geologic processes [6], [8], [9].

In this work, we will use *stable* to refer to the family, *Lévy* to refer to the special case, and *α-stable* to refer to stable distributions that are not special cases. Extensive background and theory regarding stable distributions are available in [12] and [7], among others; we will only review the theoretical aspects of stable distributions that are necessary to understand our application.

A. Stable Distribution Background

The α-stable distribution exhibits tremendous flexibility because it is described by four parameters: $\alpha, \beta, \gamma, \mu$. These parameters and some of their properties are listed in Table I.

Special cases of the stable family include the Cauchy, Gaussian (or Normal) distribution, and the Lévy (or Pearson V) distributions. Except for these special cases, a closed-form solution for the probability density function (PDF) of an α-stable random variable (RV) $Z \sim S(\alpha, \beta, \gamma, \mu)$ does not exist. Z is instead defined through its characteristic function [12]

$$E\left[e^{i\theta Z}\right] = e^{-\gamma^{\alpha}|\theta|^{\alpha}\left[1-i\beta\tan(\frac{\pi\alpha}{2})\mathbf{sign}(\theta)\right]+i\mu\theta} \quad (1)$$

for the case of $\alpha \neq 1$, and

$$E\left[e^{i\theta Z}\right] = e^{-\gamma|\theta|\left[1+i\beta\frac{2}{\pi}(ln|\theta|)\,\mathbf{sign}(\theta)\right]+i\mu\theta} \quad (2)$$

for $\alpha = 1$.

Due to this lack of a closed-form, α-stable models require additional cost to implement. We believe, however, that the potential gain in approximation accuracy offsets the increased model complexity, particularly if computational costs can be mitigated in some manner. Since network traffic is heavy-tailed and non-Gaussian, particularly at small aggregations, we will apply the positive-only (wholly-skewed) case of the α-stable distribution in our network traffic model and detector.

TABLE I: Parameters of the α-Stable Distribution

Parameter	Property	Range
α	Tail size	$(0, 2]$
β	Asymmetry	$[-1, 1]$
γ	Spread	$[0, \infty)$
μ	Location	$\Re$

B. The PαS Distribution

The PαS distribution is a wholly-positive parameterization of the stable distribution, obtained for constrained cases of ($\beta = 1, \alpha < 1$, appropriate μ) [12], which makes the PαS distribution particularly suited to describing heavy-tailed (e.g., high-variance) random processes that do not have negative components [6]. Random processes including radar clutter, edge detection in images, molecular vibration, animal foraging, and UAV search patterns and detection avoidance have been accurately modeled using the PαS or Lévy distributions.

Even when confined to the positive half of the number line, the PαS distribution has three degrees of freedom available to describe tail size, spread, and location. This freedom should enable a more precise fit to a variety of traffic features, as demonstrated in Figs. 1 and 2.

These figures show the MATLAB [13] maximum likelihood (ML) α-stable fit of two data series from the ISCX [14] data set, and the corresponding ML Gaussian fit. The relatively poor Gaussian fits are due to outliers resulting from "noisy" packet counts due to collecting small windows of data on a relatively low-traffic link in the case of Fig. 1 (the reference case for this paper), or the onset of a denial of service attack (Fig. 2). In both cases, the heavy tails and descriptivity of the PαS distribution permit more accurate modeling of the marginal feature distribution (packets per counting period) than the Gaussian distribution, whose log-likelihood fits are given for comparison in Table II. (Note that the exponential values are provided to enable comparison with other types of data series that may have a large number of zero-event counting periods, vice suggesting viability as a network traffic model.)

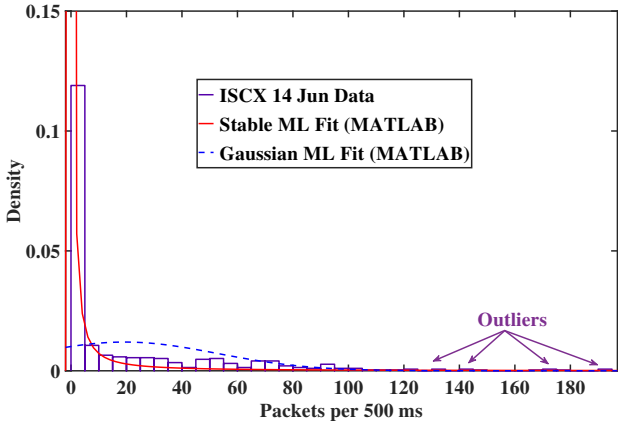

Fig. 1: ML fit comparison for impulsive 14 Jun ISCX data using a five minute data window. Fit results are $f_{\alpha} \sim S(0.40, 1, 0.30, 0.19)$ and $f_N \sim N(19.3, 1108.1)$.

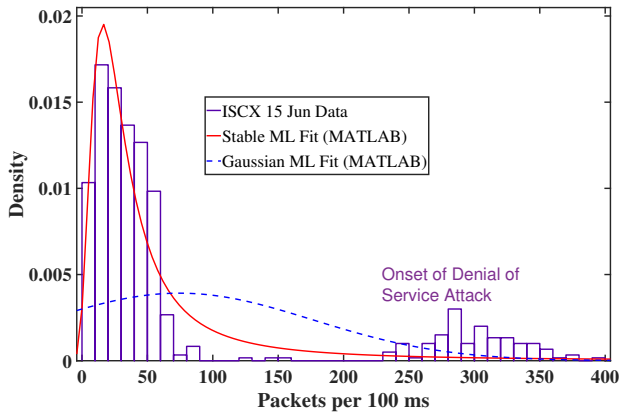

Fig. 2: ML fit comparison for 15 Jun ISCX data during attack onset using a one minute data window. Fit results are $f_\alpha \sim S(0.98, 1, 13.9, 23.3)$ and $f_N \sim N(66.5, 8781)$.

TABLE II: Log-likelihood of ML Fits of Selected Distributions to Two Network Traffic Scenarios

Traffic Scenario	α-Stable	Gaussian	Exponential
14 Jun (Noisy)	-1582	-2880	-2316
15 Jun (DoS)	-3027	-3575	-3118

Recent work has proposed a closed-form solution for the PαS case using Fox's H-function and Meijer's G-function, and applied these functions to the problem of CFAR radar detection in PαS clutter [9]. However, because this approach does not appear to support real-time implementation for our problem, we have developed previous work by Kuruoglu that accurately approximates PαS distributions through a weighted mixture of Lévy distributions [10].

IV. Lévy Approximation of PαS Distributions

To balance the competing objectives of computational tractability and accurate estimation under non-Gaussian scenarios, α-stable approximations have been developed. These approximations use mixtures of α-stable special cases, including Gaussian, Cauchy-Gaussian, and Bi-parameter Cauchy-Gaussian to leverage the closed-form expressions of their components while providing better fits [15].

A. Theoretical Foundations

Similar to these mixture methods, a PαS random variable (RV) can be approximated using a weighted mixture of Lévy RVs [10]. This approach is advantageous because in the Lévy special case, $\alpha = 0.5$ and the PDF of X can be defined in closed form as

$$f_X(x) = \sqrt{\frac{\gamma_X}{2\pi}} \frac{e^{-\gamma_X/2x}}{x^{3/2}} \tag{3}$$

The PαS random variable Z can be decomposed [10] using

$$Z = XY^{1/\alpha_X} \tag{4}$$

where Y is the mixing function, another α-stable RV, and X is distributed per (3).

The PDF of Z can then be discretely approximated [10] by

$$\hat{f}_Z(z) = \frac{1}{C} \sum_{i=1}^{N} \left(\frac{f_Y(y_i)}{y_i^3} f_X\left(\frac{z}{{y_i}^2}\right) \right) \tag{5}$$

where

$$\gamma_Y = \frac{\gamma_Z}{2^{\alpha_Z} \gamma_X^{2\alpha_Z}} \frac{\cos(\pi\alpha_Z)}{\cos(\pi\alpha_Z/2)}, \tag{6}$$

$$\alpha_Y = \alpha_Z/2, \tag{7}$$

C is a constant, i is an integer $\in [1, N]$, and N is the number of sampling points (as well as mixture components).

Examining (5), the left terms inside the summation are weighting constants, determined by choosing sample points y_i while the right terms are Lévy RVs. Both terms are scaled using the sample points and finally normalized using C.

Note that (5) applies to the case where $\alpha_Z \in (0, 0.5)$; similar expressions exist for $\alpha_Z \in (0.5, 1)$ [10]. This work only considers the case of $\alpha_Z \in (0, 0.5)$.

B. Levy Mixture Approximation (LMA)

To develop our LMA algorithm and formula for C, we extended the relationships and theory in [10] using similar approaches in the literature, such as for mixtures of Gaussians [15].

As an input, the LMA algorithm requires the parameters of a PαS RV Z with $\alpha_z < 0.5$. From Z, the mixing function $f_Y(y)$ is then generated, which is used to obtain N weights and scaling factors y_i and $f_Y(y_i)$. These weights and scaling factors are in turn applied to N Lévy functions; the weighted sum of these *components* approximates the original distribution of Z. This process is summarized in Algorithm 1 and shown graphically in Fig. 3.

To determine the relative accuracy of the resulting LMA, we measure approximation error using the Hellinger distance

$$d_{Hel}\Big(f_Z(z), \hat{f}_Z(z)\Big) = \left(\frac{2}{M} \sum_{j=1}^{M} \left(\sqrt{f_Z(z_j)} - \sqrt{\hat{f}_Z(z_j)} \right)^2 \right)^{1/2} \tag{8}$$

where $f_Z(z_j)$ and $\hat{f}_Z(z_j)$ are the reference distribution and LMA respectively, sampled at points $j \in [1, M]$.

Algorithm 1 LMA of PαS Data Series, $\alpha \in (0, 0.5)$

Inputs: Data and N
Output: LMA of $f_Z(z)$
1: Fit $f_Z(z)$ to data, confirm $\alpha_Z < 0.5$
2: Compute α_Y,γ_Y from α_Z,γ_Z using (6),(7),$\gamma_X = 1$
3: Generate mixing PDF $f_Y(y)$
4: Assign sample points y_i for $i \in [1, N]$
5: **for** y_i **do**
6: Compute $f_Y(y_i)$
7: Generate scaled Lévy component $f_X(z/y_i^2)$
8: **end for**
9: Compute $C = \sum_N \left(f_Y(y_i)/(2y_i)\right)$
10: **return** $\hat{f}_Z(z)$ using (5)

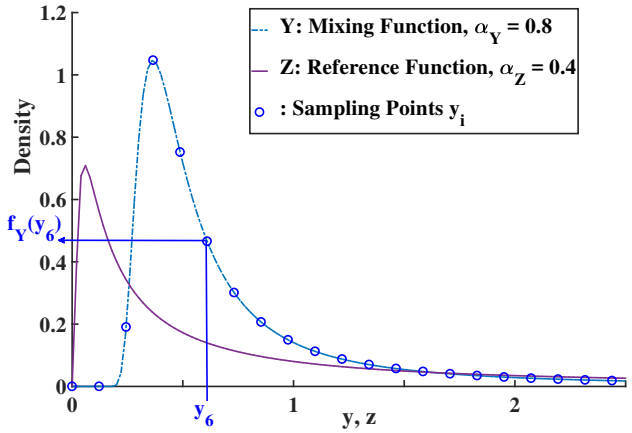

Fig. 3: Method of obtaining Lévy mixture weights from mixing function for sample point $i = 6$ and α_Z = 0.4 for the hypothetical case of $f_Z(z) \sim S(0.4, 1, 1, 0)$.

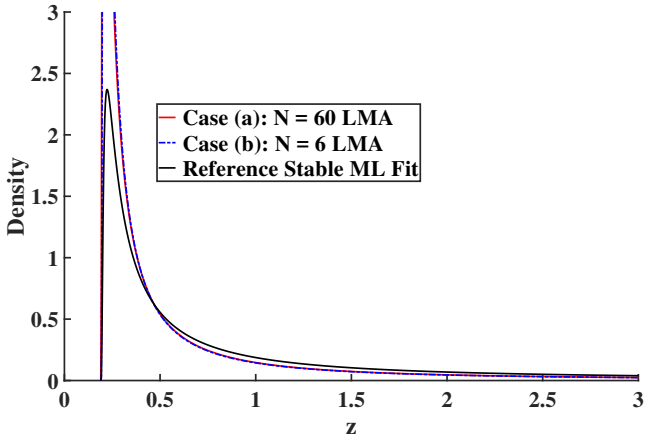

Fig. 4: Comparison of reference ML stable distribution and LMA for $N = 60$ and $N = 6$ cases with uniformly-distributed points. Case (a) $d_{Hel} = 5.09e-2$ and Case (b) $d_{Hel} = 5.00e-2$.

We use average Hellinger distance as a measure of the *absolute* error of fit because (8) does not generate positive and negative terms which can offset accumulated error during summation, as compared to Kullback-Leibler divergence. Our formula in (8) is a normalized version of the standard Hellinger distance to facilitate comparison between scenarios [16].

MATLAB's built-in maximum likelihood (ML) fitting and histogram functions were used to generate distribution estimations and results [13].

V. Results

Gaussian mixture approaches recommend uniform sampling of the mixing function at a large number of points N, or using a post-processing algorithm to optimize the sample point placement [15]. Our results demonstrate that, at least for the PαS method, approximation error is relatively independent of N and largely dominated by sample point placement. It then becomes possible to improve computational efficiency and overall accuracy through selectively applying sample points, using either linear sampling around the peak of the reference distribution, or a lightweight sample point placement algorithm.

A. Importance of Sample Point Location

The approximation results for two values of N are shown in Fig. 4, which demonstrates that a large number of sample points is not required. Case (b) was obtained using only $N = 6$ components and is more accurate than Case (a) in terms of average Hellinger distance. The fits of both cases are visually indistinguishable at the scale shown in Fig. 4.

Fig. 4 also demonstrates that sample point placement is more important than the number of points in determining LMA accuracy. Case (a) samples were uniformly distributed $\in [0, 1]$ while Case (b) samples were uniformly distributed $\in [0, 1.2]$. The slight reduction in fit error while using fewer points was due to shifting distribution bounds and the resultant change in sample point location.

B. Sampling Location Optimization

The previous result encourages exploration of optimal sample point placement routines. We have found that when using small *N*, non-uniform sampling of the mixing function can be utilized to improve approximation accuracy, as shown in Fig. 5.

Based on our qualitative assessments as well as Fig. 5, the best approximation accuracy is generally obtained by concentrating sampling points around the peak and areas of maximum change of $f_Y(y)$ (*i.e.* points y_3, y_4, and y_5 in Fig. 3). We examined other methods such as sampling only the left or right tails. These alternatives produced inferior results and were omitted from Fig. 5 for display purposes.

Fig. 5 reinforces the observation from Fig. 4 that accuracy is relatively invariant to the number of components (at least as long as the sample points are located appropriately). The significant variation of the linear method fit error with $N \in [5, 10]$ is due to the changing sample locations, as seen previously.

Overall, the results in Figs. 4 and 5 show that post-

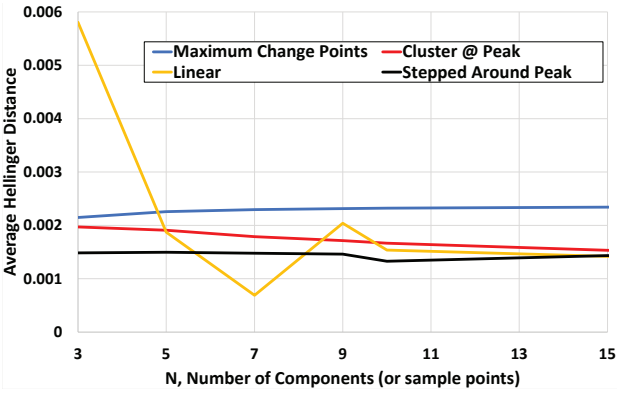

Fig. 5: Comparison of fit errors with varying sample point location schemes and N.

processing is not strictly necessarily to fine-tune the sample locations, though this added step could be used improve the relative accuracy for a given error measure, as done for the LMA in Fig. 6.

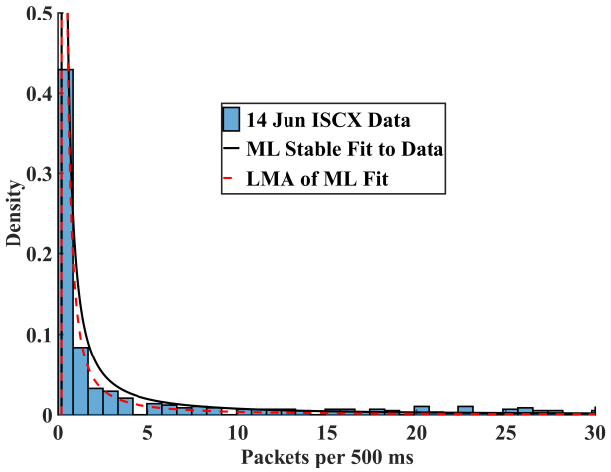

Fig. 6: Result of LMA of ML Stable fit using ($N = 4$) selectively-placed points. The data histogram is from the 14 Jun ISCX data in Fig. 1.

This figure shows the end-to-end approximation result of applying the LMA algorithm to the data in Fig. 1. In this case only four sample points were used, though their placement was manually optimized. The resulting $d_{Hel} = 4.32e - 2$, smaller than the six and 60-component LMAs.

Additional investigation of sampling optimization methods is an item of future work, and should allow quantitative evaluation of cost-accuracy trade-offs.

VI. Conclusion

In this work, we approximate a PαS RV generated from physical network traffic with reasonable accuracy using as few as four weighted Lévy distributions. Further, we have demonstrated that Lévy mixing theory [10] can be optimized to maximize approximation accuracy while controlling computational cost. The relative invariance of the fit to the number of mixture components proves the scalability of this approximation technique. Going forward, we will consider how best to apply this approximated output to a detection system.

The significance of our approach is that by using closed-form solutions to approximate an α-stable process, real-time implementations may be developed which leverage the improved fit to impulsive, skewed time series. This approach is not limited to the computer networking field but can be extended to any discipline that has computationally-efficient requirements for modeling and processing of non-Gaussian, positive (or negative) data.

Acknowledgment

The authors wish to thank Mark Kragh for his work investigating and quantifying the impact of sampling point locations, as shown in Fig. 5.

References

[1] W. Willinger, V. Paxson, and M. S. Taqqu, "Self-similarity and heavy tails: Structural modeling of network traffic," *A practical guide to heavy tails: statistical techniques and applications*, vol. 23, pp. 27–53, 1998.

[2] A. Karasaridis and D. Hatzinakos, "Network heavy traffic modeling using/spl alpha/-stable self-similar processes," *IEEE Transactions on Communications*, vol. 49, no. 7, pp. 1203–1214, 2001.

[3] A. Scherrer, N. Larrieu, P. Owezarski, P. Borgnat, and P. Abry, "Non-gaussian and long memory statistical characterizations for internet traffic with anomalies," *IEEE Transactions on Dependable and Secure Computing*, vol. 4, no. 1, pp. 56–70, 2007.

[4] F. Simmross-Wattenberg, A. Tristan-Vega, P. Casaseca-de-la Higuera, J. I. Asensio-Perez, M. Martın-Fernandez, Y. A. Dimitriadis, and C. Alberola-Lopez, "Modelling network traffic as α-stable stochastic processes: An approach towards anomaly detection," *Proc. VII Jornadas de Ingenierıa Telematica (JITEL)*, pp. 25–32, 2008.

[5] F. Simmross-Wattenberg, J. I. Asensio-Perez, P. Casaseca-de-la Higuera, M. Martin-Fernandez, I. A. Dimitriadis, and C. Alberola-Lopez, "Anomaly detection in network traffic based on statistical inference and alpha-stable modeling," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 4, pp. 494–509, 2011.

[6] R. D. Pierce, "Application of the positive alpha-stable distribution," in *Higher-Order Statistics, 1997., Proceedings of the IEEE Signal Processing Workshop on*. IEEE, 1997, pp. 420–424.

[7] J. P. Nolan, *Stable Distributions - Models for Heavy Tailed Data*. Boston: Birkhauser, 2017, in progress, Chapter 1 online at http://fs2.american.edu/jpnolan/www/stable/stable.html.

[8] R. Feldman and M. Taqqu, *A practical guide to heavy tails: statistical techniques and applications*. Springer Science & Business Media, 1998.

[9] V. A. Aalo, K. P. Peppas, and G. Efthymoglou, "Performance of ca-cfar detectors in nonhomogeneous positive alpha-stable clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 3, pp. 2027–2038, 2015.

[10] E. E. Kuruoglu, "Analytical representation for positive/spl alpha/-stable densities," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, vol. 6. IEEE, 2003, pp. VI–729.

[11] V. Paxson and S. Floyd, "Wide area traffic: the failure of poisson modeling," *IEEE/ACM Transactions on Networking (ToN)*, vol. 3, no. 3, pp. 226–244, 1995.

[12] G. Samoradnitsky and M. S. Taqqu, *Stable non-Gaussian random processes: stochastic models with infinite variance*. CRC press, 1994, vol. 1.

[13] MATLAB, *version 9.2 (R2017a)*. Natick, Massachusetts: The MathWorks Inc., 2017.

[14] A. Shiravi, H. Shiravi, M. Tavallaee, and A. A. Ghorbani, "Toward developing a systematic approach to generate benchmark datasets for intrusion detection," *computers & security*, vol. 31, no. 3, pp. 357–374, 2012.

[15] E. E. Kuruoglu, C. Molina, and W. J. Fitzgerald, "Approximation of α-stable probability densities using finite gaussian mixtures," in *Signal Processing Conference (EUSIPCO 1998), 9th European*. IEEE, 1998, pp. 1–4.

[16] D. J. Weller-Fahy, B. J. Borghetti, and A. A. Sodemann, "A survey of distance and similarity measures used within network intrusion anomaly detection," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 70–91, 2015.